

# Validation of reference genes in leaf-cutting ant *Atta sexdens rubropilosa* in different developmental stages and tissues

Ariele C. Moreira<sup>1</sup>, Adriana M. dos Santos<sup>1</sup>, Renato Lajarim Carneiro<sup>1</sup>, Odair C. Bueno<sup>2</sup>, Dulce Helena F. Souza<sup>1\*</sup>

<sup>1</sup>Department of Chemistry, Federal University of Sao Carlos, São Carlos-SP, Brazil

\*dulce@ufscar.br

<sup>2</sup>Center for the Study of Social Insects, UNESP - São Paulo State University, Rio Claro, SP, Brazil

**Abstract**— *Atta sexdens rubropilosa* is an important leaf-cutting ant species considered as a pest in agricultural crop or reforestation areas. Quantitative real-time reverse transcriptase-polymerase chain reaction (RT-qPCR) is a technique that can help us to understand the regulation and the function of a gene. However, its reliability depend on the data normalization. Different normalization strategies can be adopted for qPCR, reference genes has been cited as one of the most effective methods. It has not been identified a universal reference for all organism and experiment. In this way, the validation of reference gene is crucial step. This is the first study to evaluate reference genes for leaf-cutting ants. To this, we analyzed the expression levels of candidate reference genes (*act*, *ef1-alpha*, *ef1-beta*, *GAPDH* and *rpl18*) in different developmental stages (larva, pupa and worker) and tissues (head, mesosoma and worker without gaster) of *A. sexdens rubropilosa*. Four different algorithms (BestKeeper, geNorm, NormFinder and comparative  $\Delta C_t$  method) were used in statistical analysis of the stability of the genes and RefFinder was used to propose a consensus list for ranking the reference genes. Our results showed that the most suitable combinations of reference gene candidates were *rpl18* and *ef1-alpha* for the different developmental stages and *rpl18* and *ef1-beta* for the different tissues. In this work, we also report the obtaining from a putative acetylcholinesterase from *A. sexdens rubropilosa* (GenBank KY464935), which was used as a target gene to confirm the reliability of reference genes suggested.

**Keywords**— Acetylcholinesterase, *Atta sexdens rubropilosa*, Developmental stages, Reference gene, RT-qPCR, Tissues.

## I. INTRODUCTION

Insects are the dominant animals in most terrestrial ecosystems, both in number of species and biomass. Among them, social insects present colonies with large

numbers of individuals and, consequently, greater biomass [1]. The diversity of ant species indicates that they rank among the most successful insects. It is estimated that 40.000 ant species exist in the world, of which about 16.000 species and subspecies have already been formally described [2]. All ants are considered eusocial. The ecological significance of ants is indisputable; however, as mankind changes the environment for agricultural or forestry development or for the construction of cities, the environment becomes less complex and there is a decrease in biodiversity; although, on the other hand, opportunistic animals (generalists) are favored [3]. Among them are some species of ants, which increase in population density and can adversely affect human activities [4, 5].

Although only few ant species are considered pests (less than 1% of the known species), the economic losses caused by them can be large, especially considering those that occur in silviculture and agriculture, both in the production and storage of food [6]. Among the economically important ant species in Brazil, leaf-cutting ants stand out. They are distributed throughout the Americas and cause major damage, particularly in South America [4, 5]. Leaf-cutting ants are the main herbivores present in the Neotropics and are also considered as pests in agricultural crop or reforestation areas [7].

In addition to the losses that leaf-cutting ants cause to agriculture, silviculture, and pastures, there are the environmental problems and poisoning of other animals, including humans, caused by excessive use of pesticides in the attempt to control these ants. One approach for the development of new ways to control this problem, while minimizing the damage to the environment, is causing the silencing of a specific gene.

Reverse transcription - quantitative real-time polymerase chain reaction (RT-qPCR) is a technique that can help us to understand the regulation and the function of a gene. Different normalization strategies can be adopted for

qPCR; however, the use of a reference gene has been cited as one the most effective methods, since the reference gene undergoes the same steps as the target gene, correcting errors and differences in the sample [8, 9]. Several works have demonstrated the importance of choosing a proper reference gene and impact of using such not appropriated gene. Incorrect results might be obtained due to misinterpretation of RNA transcription levels especially for low abundance gene transcripts [10, 11]. To date, several works were developed to determine the reference gene in Insecta, which demonstrate that is impossible to find a universal reference gene able to covering all organism and conditions [11, 12, 13, 14, 15, 16, 17].

The choice of a gene as reference gene is not trivial and starts with the selection of candidate reference genes to be analyzed [18]. Housekeeping genes (HKG) are usually first selected to be investigated as reference genes due to the assumption that they are involved in essential processes for the survival of cells and are expected to be expressed in a stable and nonregulated level [19]. A reliable reference gene should exhibit an expression level not affected by experimental factors, with minimal variability between tissues and physiological states and a Ct (Cycle Threshold) similar to the target gene [9]. The most studied reference genes, GAPDH and 18s rRNA, are not always expressed in a constant manner. In addition, their expression can be altered depending on the organisms and their life stages [9, 12].

A good strategy for selecting potential candidate reference genes is based on previous data from species relative to the studied specie due to the high degree of similarity between genomes and the expectation of a similar expression level [9]. However, differences in stability have been verified in the analysis of a reference gene in Insecta for organisms from the same order [11, 12, 14, 15], family [20, 21] and even for those of the same genus [14, 16, 21]. This justifies studies to validate reference genes for an organism and experimental conditions before the analysis for precise mRNA quantification [22].

We believe that this is the first study to evaluate reference genes for leaf-cutting ants. Genome from *Acromyrmex echinator* [23] and *Atta cephalotes* [24], both leaf-cutting ants, are available in the database but there is no validated gene(s) for leaf-cutting ants. In the present study, seven candidate reference genes from *A. sexdens rubropilosa* were selected: actin (*act*), elongation factor 1-alpha (*ef1-alpha*), elongation factor 1-beta (*ef1-beta*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), ribosomal protein L18 (*rpl18*), TATA box binding protein (*tbp*), and 18S ribosomal RNA (*18S rRNA*). The expression level and stability of them, except *tbp* and *18S*

*rRNA* were examined. The stability of these candidates was investigated in three *A. sexdens rubropilosa* developmental stages (larva, pupa and worker) and in the following parts of the insect: head, mesosoma, gaster, and worker without gaster.

To validate the results the expression profile of a putative target gene, acetylcholinesterase from *A. sexdens rubropilosa*, was investigated. Acetylcholinesterase (AChE, EC 3.1.1.7) is a serino hydrolase that hydrolyzes and inactivates the neurotransmitter acetylcholine controlling the cholinergic signal transmission in the synapse [25]. The evaluation with the target gene emphasize the importance to validate the reference gene as internal control in genomic research and the results presented will be useful for further works in this field for leaf-cutting ants.

## II. MATERIAL AND METHODS

### 2.1 Biological samples

The *A. sexdens rubropilosa* Forel (Hymenoptera: Formicidae) was collected from laboratory nest localized in the Center of Studies on Social Insects (UNESP, Rio Claro, Brazil). The nest was supplied daily with leaves of *Eucalyptus alba*, oat seeds and occasionally with the leaves of other plants such as *Hibiscus* sp., *Ligustrum* sp. or rosebush petals.

Developmental stages samples were picked from the nest: 10 larvae, 10 pupae and 10 workers were collected for each replicate, washed with RNase-free phosphate-buffered saline (PBS) and stored at -80 °C until used. Tissue samples were dissected from workers: 10 heads, 10 mesosomata, 10 gasters and 10 workers without gaster for each replicate, followed by wash with PBS and stored at -80 °C until RNA extraction. All samples were collected in triplicate (biological triplicate).

### 2.2 RNA extraction and cDNA synthesis

Total RNA from larvae, pupae and workers without gaster was extracted using a combined method with TRIzol® (Thermo Fisher Scientific) and PureLink® RNA mini Kit (Thermo Fisher Scientific). For tissue samples from workers, head, mesosoma and gaster, only the PureLink® kit was used. The manufacturer's protocol was followed for both applications. Total RNA from each sample was diluted in 10 mM Tris-HCl pH 7.5 and the quantity and quality of the samples were determined by the 260/280nm and 260/230nm ratio using a BioSpec-nano (Shimadzu-biotech). The RNA integrity was analyzed by agarose denaturing gel 1.2 % (w/v) and confirmed by the intense ribosomal RNA bands and the absence of smears. The total RNA was treated with DNase (DNaseI, RNase-free -Thermo Fisher Scientific) to eliminate potential genomic DNA contamination.

First-strand complementary DNA (cDNA) was synthesized using 1.35 µg of total RNA with SuperScript® VILO Master Mix (Thermo Fisher Scientific) with 20 µl final reaction volume, following instructions from the manufacturer. The synthesis of cDNA was performed in triplicate for each sample (replicate) and the product was stored at -20°C for later use.

### 2.3 Selection and procedure for obtaining the sequence of candidate reference genes and the putative acetylcholinesterase gene

Seven genes were selected as candidate to reference genes: actin (*act*), elongation factor 1-beta (*efl-beta*), elongation factor 1-alpha (*efl-alpha*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), ribosomal protein L18 (*rpl18*), TATA box binding protein (*tbp*), and 18S ribosomal RNA (*18S rRNA*). Acetylcholinesterase (AChE) was used as a target gene.

*A. sexdens rubropilosa* genome is not yet known, for this reason, the sequences of most of these genes are not yet deposited. Sequence alignments for every gene from several species of ants were performed and conserved sequence regions were used to design specific and/or degenerated forward and reverse primers (Table 1).

The DNAs were amplified by PCR performed with 1 µL cDNA; 1 µM for the specific primer and 2 µM for the degenerated one; 0.2 mM of dNTPs and 1.25 U of Pfu DNA Polymerase (Thermo Fisher Scientific) in 25 µL final volume. PCR amplification was performed using the following program: 3 min at 95 °C followed by 40 cycles of 1 min at 95 °C, 90 s at 52 or 62 °C, 6 min at 72 °C and a final extension step of 10 min at 72 °C. The amplification products were evaluated on 1% agarose gel and the bands were extracted and purified. Samples were quantified by absorbance in 260 nm and then submitted for sequencing analysis (ABI 3730 DNA Analyser - Thermo Fisher Scientific) with the same primers used for amplification. The sequences were analyzed with *BioEdit* (v7.2.5-<http://www.mbio.ncsu.edu/bioedit/page2.html>) and the search for similarity was carried out using the *BLAST tool*. The amplicons were compared with ants' sequences and led to an identity of over 90 %.

Acetylcholinesterase sequence from *Acromyrmex echinator* (GenBank GL888116.1) was used to design primers with the inclusion of site for restriction enzymes and exclusion of signal peptide (Table 1). The PCR reaction was performed similar as described above with 0.2 µM of each primer, an annealing temperature of 63°C in 30 cycles with an extension time of 3 min. The reaction product was analyzed on 1% agarose gel, purified and sequenced.

### 2.4 Primer design for quantitative real-time RT-PCR (RT-qPCR)

Using the sequence from the amplicons, new primer pairs were designed, for each gene, using Primer Express® Software Version 3.0 and selecting the amplicon length between 50 and 150 pb. Among the various possibilities of primers provided by the software output, the selection of the primer pair was based on the low score penalty and smaller size of the amplicon. Primer sequences and amplicon characteristics are summarized in Table 2 for each candidate gene and for the AChE gene.

### 2.5 Quantitative Real-time PCR (qPCR)

The minimal primer concentration was determined using two-by-two combinations of forward and reverse primers in 100, 150 and 300 nM, in duplicate, and a non-template control for each combination. RT-qPCR was performed in an Applied Biosystem StepOnePlus™ system (Thermo Fisher Scientific) with a total reaction volume of 12 µL, containing 6 µL *Power SYBR® Green PCR Master Mix* (Thermo Fisher Scientific), 3 µL of forward and reverse primers in the appropriate concentration to give the relation described above and 3 µL of cDNA previously 30-fold diluted. Cycling conditions were: 10 min at 95 °C (polymerase activation) followed by 40 cycles at 95 °C during 10 s (denaturation) and 60 °C during 1 min (annealing/extension). For each reaction, the dissociation of the PCR products (melting curve) was analyzed from 60 to 95 °C to ensure the specificity of the amplified product.

The appropriate primer concentration was used to determine the RT-qPCR efficiency by a relative standard curve for each candidate reference genes. For this purpose, a 5-fold serial dilution of the cDNA was used as a template molecule for candidate reference genes and a 2-fold serial dilution for acetylcholinesterase. Samples were analyzed in triplicate plus a negative control. RT-qPCR efficiency was calculated according to the equation 1, in which the slope comes from the plot of Ct values against the logarithm of cDNA concentration [10]. Efficiencies between 90% and 110% were used for further statistical analysis (Table 2).

$$E = (10^{-1/\text{slope}} - 1) \times 100 \quad (1)$$

Once the optimum primers and cDNA concentrations were determined for each gene, the gene expression analysis was performed by RT-qPCR in the conditions already described above using 3 µL of cDNA diluted 60 times. The reaction was performed in triplicate and with a non-template control for each conversion reaction of cDNA.

### 2.6 Data Analysis and Statistics

### 2.6.1 Expression level analysis

The expression level of candidates for reference gene was analyzed by standard deviation, coefficient variation and Student's *t*-test. The *t*-test was used to verify if the mean value of expression levels between two different stages of development are statistically different or not. The same procedure was adopted to analyze if there is significant difference among the expression levels between different parts of the ant body. For this comparison, the first procedure is to calculate the pooled estimate of standard deviation (2), followed by the calculation of the experimental *t*-value (3), where *S*, *n* and  $\bar{x}$  are the standard deviation, degrees of freedom and means, respectively, for the two analyzed genes. If the experimental *t*-value is lower than the critical *t*-value then there is no significant difference between the mean values of the gene expression at a 95% of confidence level [26].

$$S_p = \sqrt{\frac{S_1^2(n_1+1) + S_2^2(n_2+1)}{n_1+n_2-2}} \quad S_p = \sqrt{\frac{S_1^2(n_1+1) + S_2^2(n_2+1)}{n_1+n_2-2}} \quad (2)$$

$$t_{exp} = \frac{|\bar{x}_1 - \bar{x}_2|}{S_p} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad t_{exp} = \frac{|\bar{x}_1 - \bar{x}_2|}{S_p} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad (3)$$

In addition, the source of the standard deviation was compared in the analyses by comparing the variability of data from replicates with data related to the different body parts or different life cycle stages.

The autoscaling preprocess also was used in order to obtain a better visualization of the most similar variables. This strategy is well known in chemometrics to normalize all the variables (expression levels, in this case) in order to minimize the differences in the intensity among them [27]. This preprocess is performed by subtracting the mean value of the genes expression in each developmental stage from the total mean of the same gene for all developmental stages, followed by division by the standard deviation of the gene for all developmental stages.

All of the tests were performed at 95% confidence level using the Microsoft Excel® software.

The relative expression level of the target gene was obtained according to the relative quantification by  $2^{-\Delta\Delta Ct}$  method [28]. For developmental stage larva was used as calibrator and mesosoma for tissue. The data was plot as mean  $\pm$  SEM and the analysis with 95% confidence using GraphPad Prism 5.

### 2.6.2 Selection of reference genes

The selection of the reference gene was performed using four algorithms, BestKeeper® version 1 [29], geNorm version 3 [30], NormFinder version v0.953 [18] and the comparative  $\Delta Ct$  method [31]. RefFinder was used to compare and rank the reference gene candidates [32]. In

addition, the results from these software were compared with the statistical analysis performed.

### 2.6.3 BestKeeper

BestKeeper is an Excel based spreadsheet software that uses raw data (Ct values) and reaction efficiency (E) to identify the best-suited standards and combines them into an index [29]. The output Table shows descriptive statistics for each reference gene candidate: the geometric mean (geo Mean), arithmetic mean (ar Mean), minimal (min) and maximal (max) value, standard deviation (SD), and coefficient of variation (CV). The x-fold over- or under-expression of individual samples are calculated based on the geometric mean. These results are corrected via RT-qPCR efficiency to exhibit minimal and maximal values considering the x-fold ratio and their SD (SD  $\pm$  x-fold). The stability of the reference gene candidate can be evaluated by the user considering the calculated variation, such as SD and CV. Reference genes can be ordered from the most stable (lowest variation) to the least stable (highest variation). Candidate genes with SD  $[\pm Ct]$  higher than 1 (= starting template variation by a factor of 2) can be considered inconsistent and it is recommended to exclude them from the calculation index [29].

BestKeeper also tests individual samples for their integrity. To do this, x-fold values are used through an intrinsic variation (InVar) for a single sample. It has been suggested that samples with 3-fold over- or under-expression should be removed from the analysis due to high deviation that can be attributed to inefficient sample preparation, incomplete reverse transcription or sample degradation [29].

### 2.6.4 geNorm

geNorm uses relative quantification data from raw Ct values by  $2^{-\Delta Ct}$ . This software determines the expression stability of candidate reference genes based on the gene-stability measure [30]. The internal control gene stability measure M is defined as the average pairwise variation for that gene with all other tested reference genes, where the lowest value for M corresponds to the most stable candidate, and the highest corresponds to the least stable. Values that surpass the cutoff value of 1.5 are not considered stable. The program enables stepwise exclusion of the gene with the highest value of M and recalculation of M for the remaining genes ranking them according their expression stability.

The second important parameter calculated by geNorm is the pairwise variation ( $V_n/V_{n+1}$ ) between two sequential normalization factors ( $NF_n$  and  $NF_{n+1}$ ) to obtain the minimal number of reference genes [30]. The cutoff value of 0.15 indicates that no additional gene, beyond the *n*



most stable genes, needs to be included for a reliable analysis.

### 2.6.5 NormFinder

NormFinder, a model-based approach for the estimation of expression variation. It is able to identify stably expressed genes in a set of reference gene candidates. The Ct values were transformed to a linear scale by the same method used for geNorm in order to prepare input data. The mathematical model of gene expression presented on a Visual Basic application for Microsoft Excel estimates the intragroup variation as well as the intergroup variation in all groups [18]. These variations are combined into a stability value, representing a practical measure of the systematic error that will be introduced when using the investigated gene. The software requirements are 8 samples/groups and at least 3 candidate reference genes; 5-10 candidates are recommended in order to obtain reliable results. The best reference gene candidates are ranked in an index that is based on stability values; a low stability value indicates the most stably expressed gene.

### 2.6.6 Comparative $\Delta C_t$ method

This method compares the relative expression of “pairs of genes” within each sample to identify a useful reference gene [31]. The variation  $\Delta C_t$  for each two genes is obtained by the difference of Ct values. The mean, standard deviation and mean of the standard deviation related to the  $\Delta C_t$  are obtained and used to rank genes. Two genes are stably expressed or co-regulated if a constant  $\Delta C_t$  value is observed between two genes. A low deviation value shows a more stable expression due to a short variability.

### 2.6.7 RefFinder

RefFinder is a web-based tool (<http://fulxie.0fees.us/?type=reference>) that considers the four algorithms described before to rank the candidate reference genes. It uses raw Ct as input data to obtain the rank provided by each program. Then, based on the ranks, a weight for each individual gene is calculated to obtain the final overall rank [32].

Table.1: Primer pairs sequence to identify the sequence of candidate reference genes.

Genes	Function	Primer sequence <sup>a</sup> (5'-3')	Amplicon size sequenced (bp)
<i>act</i>	Cytoskeletal structural protein involved in cell motility, structure and integrity	F: GYGACGACGAMGTAGC R: TGCCAGATCTTCTCC	259
<i>ef1-alpha</i>	Elongation during polypeptide synthesis in the ribosome	F: GACATTGCCTTGTGGAAG R: CAGTTGGCCTGGTAGGTGGC	498
<i>ef1-beta</i>	Elongation during polypeptide synthesis in the ribosome	F: GTGGCAACCAACTCAGG R: GTGGACGAAGCTGGG	177
<i>GAPDH</i>	Carbohydrate metabolism	F: CAACTTYGARRTYSTCGAGG R: CCRWAYTCGTTGTCATACC	436
<i>rpl18</i>	Encode a ribosomal protein that is a component of the 60S subunit	F: CGATATTAATCATAAGCATGATCG GA R: CTTATAACCGCAGCTGCGTC	481
<i>tbp</i>	Coordinate the initiation of transcription by RNA polymerase II promoter	F: ATGGATCAGATGCTTCCG R: AGACCTGGAAATAGCTCTGG	677
<i>18S rRNA</i>	Structural RNA constituent of subunit 40S of the ribosome	F: AGCCATGCATGTCTCAGTGC R: CGCGACGGGATATTAGTTGG	648

<sup>a</sup> F and R indicate forward and reverse primers, respectively.

Table 2: Primer sequences and amplicon characteristics for reference gene candidates used in RT-qPCR analysis.

Gene	Sequence (5'-3')	Product Length (bp)	Efficiency (%) <sup>b</sup>	R <sup>2c</sup>
<i>act</i>	F: TCCTCGCGCCGTCTTTC R: TTGACCCATACCGACCATCA	69	98.2	0.990

<i>ef1-alpha</i>	F: AGCCGCTGTTGCATTCGT			
	R: TGACGGATACTTCCAACATATTGTC	64	95.1	0.993
<i>ef1-beta</i>	F: GGCAACCAACTCAGGCTGAT			
	R: CAACGGAGTACATGAGGATTCTG	82	99.8	0.900
<i>GAPDH</i>	F: ATGACGACTGTACATGCGATTACA			
	R: TCACGCCATAGCTTGCTTGA	70	97.4	0.990
<i>rpl18</i>	F: CGAGATCATCACGTTTCGATC			
	R: CTGCATCAAGACTGTACGTTTTCC	66	97.9	0.988
<i>tbp</i>	F: CAGCAGTCACAACAATTTCAACAA			
	R: TCATTAGCATGCCACTCTGCAT	75	*	
<i>18S rRNA</i>	F: CTGATCGCACGGTCTTAGCA			
	R: CAGAACCTACCATCGACAGTTGAT	73	*	

<sup>a</sup>F and R indicate forward and reverse primers, respectively

<sup>b</sup>RT-qPCR efficiency, calculated by the standard curve method

<sup>c</sup>Determination coefficient

\*Results will be discussed in item 3.3

### III. RESULTS

#### 3.1 Sample quality

Despite the accurate validation of reference genes, several problems can directly influence the results during the sample processing and preparation. In general, these problems can be associated to factors such as sample storage, RNA extraction and quality, synthesis of cDNA with transcriptase reverse, primer design and normalization [33]. Agarose gel electrophoresis was used to confirm the integrity of the RNA extracted from *A. sexdens rubropilosa* (Fig. 1). As previously described for insects [34], only one intense RNA band can be seen in the denaturing gel, which corresponds to the two fragments of the 28S rRNA that co-migrate with 18S rRNA.

The RNA extracted from worker and from gaster using the Trizol method was degraded. Valles and collaborators detected the presence of an endogenous component located in the abdomen of adult ants (terminal abdominal segments) from *Nylanderia pubens* Forel (Hymenoptera: Formicidae), and also in queens and alate ants, capable of degrading RNA [35]. This report has also showed that the addition of at least 50 mM EDTA leads to intact RNA. However, EDTA can inhibit subsequent transcription and the PCR reaction, which could include one more variable in RT-qPCR experiments [36]. Therefore, new RNA extraction was carried out with the PureLink® RNA mini Kit producing intact RNA from the worker (Fig. 1, lane 3) and partially intact from the worker's gaster (Fig. 1, lane 6). RNA extracted by combining Trizol with the kit (from larvae, pupae and workers without gaster) (Fig. 1, lanes 1, 2, and 7, respectively) and only with the kit (head and mesosoma) (Fig. 1, lanes 4 and 5) showed characteristic bands of intact RNA.

There are divergent discussions about the influence of RNA integrity on RT-qPCR experiments. Some authors [36] suggest that RNA degradation can be tolerated since an amplicon with 70-250 bp is obtained, while other authors indicate that partially degraded RNA can give an imprecise result of genic expression [37]. Because of this, the RNA from the ant's gaster was excluded from the analysis with exception of the BestKeeper algorithm that also analyze the sample integrity.

#### 3.2 Selection and procedure for obtaining the sequence of reference gene candidates and AChE gene

The lack of genome information for *A. sexdens rubropilosa* was not an obstacle for gene validation: the sequences alignment of other ant nucleotides and the analysis of the conserved regions enabled the design of primers, which were used for obtaining amplicons from *A. sexdens rubropilosa* cDNA (Table 1). The choice of a reference candidate for analysis was made based on the reference gene for *Solenopsis invicta*, the closest insect (Formicidae) with described reference genes [12] and other insects [9, 13, 15, 21, 38, 39, 40].

All seven candidate reference genes (*act*, *ef1-beta*, *ef1-alpha*, *GAPDH*, *rpl18*, *tbp* and *18S rRNA*) were amplified by PCR using these primer pairs. The amplicons were sequenced and these sequences were used in a sequence similarity search, confirming the identity of the genes. The AChE sequence from *A. sexdens rubropilosa* without signal peptide can be accessed in GenBank KY464935. The DNA sequence amplified from the candidate reference genes and AChE was used to design specific primers for the RT-qPCR (Table 2).

### 3.3 Standardization of the conditions for Quantitative Real-time PCR (qPCR)

The minimum primer concentration for each target gene was determined to minimize non-specific amplifications and to reach the maximum amplification efficiency [41]. The proper combinations of primers were considered those that introduced the melting curve with a single peak, resulting in amplification reactions with lower Ct values and greater  $\Delta R_n$ . All samples showed a single peak in the melting curve. However, it was also observed one peak in the melting curve of the non-template control for the *thp* gene suggesting the formation of a primer dimer. The primers pairs designed for the *thp* gene could provide unreliable results and, therefore, this gene was excluded from this study. Nevertheless, analysis would be possible through the design of new primers for this gene. The best primer concentrations determined for the remaining six genes were used in the determination of the RT-qPCR efficiency for each gene. The *18S rRNA* showed a high abundance of transcripts due to the low value of Ct (data not shown). The sample (cDNA) was diluted by a factor of 60 to verify the reaction efficiency, but the results were not satisfactory. The discrepancy between rRNA and mRNA has been discussed as a negative point in the use of rRNA in reference genes studies [30]; in addition, the necessity of high sample dilution prior to qPCR can lead on dilution errors [42]. For these reasons, studies that analyzed this gene as a reference gene in insects also suggested the elimination of *18S rRNA* from the list of consensus genes [17, 21]. Therefore, *18S rRNA* gene was excluded from this study. RT-qPCR efficiency for *act*, *efl-alpha*, *efl-beta*, *GAPDH*, *rpl18* and *AChE* was between 95.1 – 103.6 %, showing that they can be used for RT-qPCR analysis (Table 2). The relative expression level of the target was obtained by  $2^{-\Delta\Delta C_t}$  method, to this the target and reference should have amplification efficiencies approximated equal, the observation of how  $\Delta C_t$  varies with template dilution showed that the method can be used for analysis (data not shown) [28].

### 3.4 Statistical analysis

#### 3.4.1 Transcription profile of candidate reference genes

Fig. 2a shows the genes plotted as function of their gene expression average at different developmental stages. The autoscaling preprocess was used in order to obtain a better view of the correlation among the variables. Fig. 2a shows that the most correlated genes are *efl-alpha* and *efl-beta*, followed by *rpl18*, being the genes *act* and *GAPDH* more intercorrelated. Therefore, the variables *efl-alpha*, *efl-beta*, and *rpl18*, are the most correlated variables and they present the low variability with the

development stage, making these variables good candidates for reference genes.

Fig. 2b shows the genes plotted as function of their gene expression average in different parts of the body using autoscaling, as explained before. As showed in this Fig., the most correlated genes with respect to different body parts are *rpl18* and *efl-beta*, followed by *GAPDH* and *efl-alpha*. In this case, although the genes *GAPDH* and *act* present the best SD and CV values (data not shown), they do not present good correlation compared to the other variables. Then, the best choice for a reference gene will depend if the algorithm used seeks lower SD and CV values or the two most correlated variables.

The Student's t-test showed no significant difference for all genes, with a confidence level of 95%, when larvae and pupae were compared. However, there were significant differences in the expression of the genes when larvae and pupae were compared to workers. *efl-alpha*, *efl-beta* and *rpl18* presented the most constant expression with the development stage. Similar results were obtained for the different tissues, where there were not significant differences for all genes expression levels when head and torax were compared. In addition, *act* was the only gene that didn't present significant differences comparing any tissue by t-test at 95% of confidence. Again, the best selected gene will depend of the algorithm used for the genes evaluation.

The comparison between the standard deviation for replicates from RT-qPCR experiment and replicates of converting RNA into cDNA showed that the deviation of the last one is 1.6 and 3.6 times higher than the first. This was expected since it is well known that the conversion of RNA into cDNA is the main source of data variability.

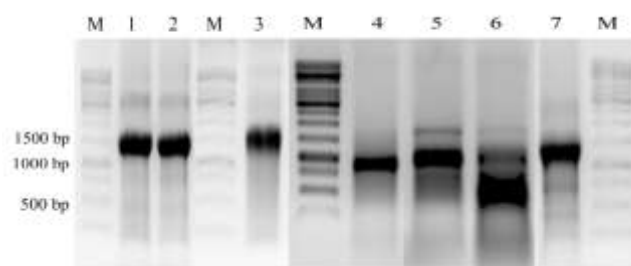


Fig. 1: Analyzes of RNA integrity by denaturing agarose gel electrophoresis 1.2% (w/v). (M) molecular marker; RNA extracted from: 1) larvae; 2) pupae; 3) workers; 4) head; 5) mesosoma; 6) gaster; 7) worker without gaster.

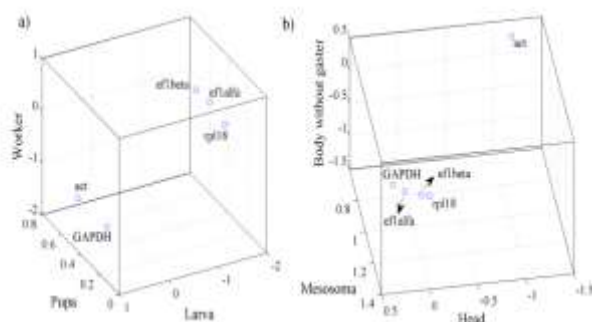


Fig. 2: Averages of gene expression as function of the developmental stage (a) and tissue (b) for each candidate. The data was auto-scaled before the plot without gaster.

### 3.4.2 BestKeeper

Using the output values for Ct variation ( $SD \pm Ct$ ), the expression level of the candidates reference genes was analyzed and the ranking was constructed. As the  $SD \pm Ct$  are below 1 for all of the five candidates from the developmental stages (larva, pupa and worker), it means that they can be considered stably expressed (Table 3). The gene stability in decreasing order for developmental stages is: *GAPDH*, *rpl18*, *efl-alpha*, *efl-beta*, and *act*. The intrinsic variance (InVar) of expression for a single sample is below 3 and the highest value obtained was 1.03. This result confirms the integrity of total RNA extracted from specimen of every developmental stage. In addition, the results agree with the statistical analysis where was verified that *GAPDH* present the lowest standard deviation.

The tissue samples (head, mesosoma, and worker without gaster) including gaster the Ct variation ( $SD \pm Ct$ ) and up/downregulation ( $SD \pm x\text{-fold}$ ) showed higher values than 1 and 2, respectively. In this way, none of the candidate reference genes from the tissues could be used. The InVar [ $\pm x\text{-fold}$ ] values for the samples were higher than 3 for the gaster confirming RNA degradation as observed on the denaturing agarose gel (Fig. 1) and justifying the exclusion of this tissue from analysis. The gaster samples were eliminated and the data were analyzed again; this has led to acceptable values of  $SD \pm Ct$ ,  $SD \pm x\text{-fold}$  and InVar [ $\pm x\text{-fold}$ ] for all reference gene candidates. This proceeding also was adopted by Ponton and collaborators that also identified InVar [ $\pm x\text{-fold}$ ] > 3 when analyzing different treatments together from *Drosophila melanogaster* [17]. To overcome this, the authors suggested a separate analysis of the samples with different treatments.

The stability of the genes obtained by BestKeeper, in decreasing order for tissues (head, mesosoma, and worker without gaster), was *GAPDH*, *act*, *rpl18*, *efl-1beta* and *efl-alpha* (Table 3).

### 3.4.3 geNorm

Vandesompele and collaborators described this robust and innovative strategy to identify the most stably expressed control genes in a given set of tissues, and to determine the minimum number of genes required to calculate a reliable normalization factor [30]. These authors also suggested the use of at least three reference genes to increase the confidence of the analysis when the suggested number of genes is too high or the sample limited.

First geNorm calculates the gene stability measure (M); the genes presenting  $M < 1.5$  are considered stable. Here, the five candidate reference genes, considering both the development stages and different tissues, could be considered for use as reference genes.

The candidate reference genes were ranked after stepwise exclusion of the highest M value (Table 3, Fig. 3), which results in a combination of two constitutively expressed genes that exhibit the most stable expression in the tested samples. The decrease of the M value during this analysis reflects the differences in the stability of reference gene candidates associated with the highest stability of the remaining genes. In this way, it is clear that *act* and *GAPDH* present an unstable expression, represented by the decrease of the M value after removal of these genes (Fig. 3) in both groups of analyzed samples (development stages and tissues).

Therefore, the most stable genes are *efl-alpha* and *efl-beta* for the different developmental stages and *efl-beta* and *rpl18* for the different tissues. These results perfectly agree with the spatial representation of the gene expression levels presents in Fig. 2.

Vandesompele and collaborators also demonstrated the large errors associated with the use of a single gene as reference gene [30]. To obtain reliable results for gene expression analysis, geNorm determines the minimum number of genes to be used as reference genes in a particular experiment. To do that, the pairwise variation was individually determined for each gene starting with the two most stable genes ( $n=2$ ) with the sequential addition of the other least stable genes ( $V_{n/n+1}$ ). The optimum number of reference genes was determined by the levels of variation in the average reference gene stability.  $V_{2/3}$  values are below the threshold value of 0.15 (Fig. 4). Then, geNorm tool indicates that the use of only two genes, the most stable ones, is sufficient to obtain accurate results for normalization experiments in RT-qPCR analysis from the different developmental stages and tissues for *A. sexdens rubropilosa*.

### 3.4.4 NormFinder

The mathematical model of NormFinder considers the inter- and intra-group variation to estimate the gene stability and rank genes with minimal variation,



eliminating problems associated with the selection of co-regulated genes [18]. For the different developmental stages, the most stably expressed candidate gene was *rpl18* with the lowest variability value. The decreasing

order of gene stability is *rpl18*, *efl-alpha*, *efl-beta*, *GAPDH* and *act* (Table 3). For different tissues, the decreasing order was *rpl18*, *efl-beta*, *GAPDH*, *efl-alpha* and *act*.

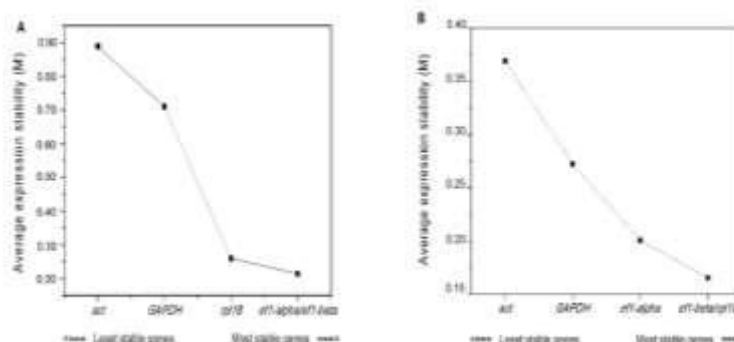


Fig.3: Gene expression analysis by geNorm. Expression stability and ranking of 5 candidate reference genes. The M value (indicates the average expression stability) is lower for the most stable expression. A) Developmental stages; B) Tissues

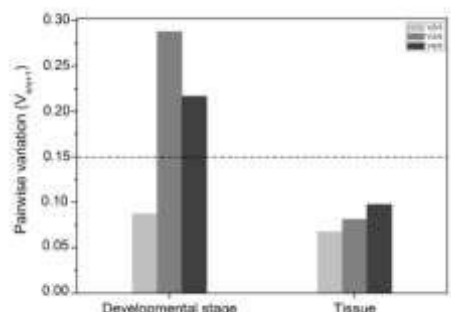


Fig. 4: Pairwise variation of candidate reference genes for determination of the optimal number of control genes for accurate normalization. Pairwise variation ( $V_{n/n+1}$ ) analysis between the normalization factors  $NF_n$  and  $NF_{n+1}$

Andersen and collaborators that elucidated the discrepancies caused by the differences between the approaches, due to the tendency of pairwise comparison to select genes with highest degree of similarity in their expression profile [18], foresaw the difference in rank obtained by NormFinder and geNorm for developmental stages. This is a problem when there are co-regulated genes between the candidates, they usually have a tendency to show very similar expression profiles and be top ranked, independently of their expression stability [18]. This can be the reason for which geNorm ranked *efl-alpha* and *efl-beta* genes as the best genes to be used as reference for different developmental stages.

### 3.4.5 The comparative $\Delta C_t$ method

The  $\Delta C_t$  method compares pairs of genes, similarly to geNorm, and uses  $\Delta C_t$  to estimate the gene variability [31]. Changes in gene variability were observed by the increase or decrease on the deviation of  $\Delta C_t$  among all possible combinations between candidate reference genes. The analysis of the most stable gene was done comparing the mean of the standard deviation of  $\Delta C_t$ . The lowest values correspond to lower variability for this gene,

which establishes it as the most stable gene. For the developmental stages, *rpl18*, *efl-alpha*, and *efl-beta* genes showed the lowest and similar deviation (Table 3). The rank in decreasing order of stability for the developmental stages was *rpl18*, *efl-alpha*, *efl-beta*, *GAPDH* and *act*. For the tissue samples the decreasing order of stability was *rpl18*, *efl-beta*, *efl-alpha*, *GAPDH* and *act*.

### 3.4.6 RefFinder

The stability of the candidate reference genes was evaluated with four different algorithms (BestKeeper, geNorm, NormFinder and comparative  $\Delta C_t$  method). Differences in the mathematical model for each one result in different ranks for gene stability, but the methods are equally important [9, 17]. RefFinder was used to propose a consensus list for ranking the reference gene for *A. sexdens rubropilosa* (Table 3), showing, for the developmental stages, in a stability decreasing order: *rpl18*, *efl-alpha*, *efl-beta*, *GAPDH* and *act*. For the tissues, the decreasing order of stability was *rpl18*, *efl-beta*, *GAPDH*, *efl-alpha* and *act*.

### 3.4.7 Expression of the AChE from *A. sexdens rubropilosa*

The RT-qPCR data for AChE normalized by each of the candidate reference genes are presented in Fig. 5a and Fig. 5b, for developmental stages and for tissues, respectively. As predicted, the results show a difference in quantification depending on the gene used to normalize. In worker the AChE expression level normalized with *act* and *GAPDH* was about 7.3-fold and about 4.2-fold lower than those normalized with *rpl18* and *efl-alpha* ( $P < 0.0001$ ). The great difference in tissues was in worker without gaster when the data were normalized with *act*, exhibiting an expression level 2-fold higher when compared with *rpl18* ( $P = 0.0012$ ). However, the statistical analysis for the two top ranked genes by RefFinder for developmental stages (*rpl18* and *efl-alpha*)

and for tissues (*rpl18* and *ef1-beta*) showed no significant difference ( $P>0.05$ ).

Table.3: Rank of reference genes in decrease order based on their expression stability according to BestKeeper, geNorm, NormFinder, comparative  $\Delta C_t$  method and RefFinder. The values were obtained after individual analysis of each software.

Developmental stage					Tissues				
BestKeeper	geNorm	NormFinder	$\Delta C_t$ method	RefFinder	BestKeeper	geNorm	NormFinder	$\Delta C_t$ method	RefFinder
GAPDH (0.50)	ef1-alpha/ef1-beta (0.21)	rpl18 (0.135)	rpl18 (0.73)	rpl18 (1.57)	GAPDH (0.33)	ef1-beta/rpl18 (0.16)	rpl18 (0.028)	rpl18 (0.30)	rpl18 (1.32)
rpl18 (0.58)	rpl18 (0.26)	ef1-alpha (0.143)	ef1-alpha (0.74)	ef1-alpha (1.86)	act (0.38)	ef1-alpha (0.20)	ef1-beta/GAPDH (0.050)	ef1-beta (0.31)	ef1-beta (2.00)
ef1-alpha (0.59)	GAPDH (0.71)	ef1-beta (0.176)	ef1-beta (0.79)	ef1-beta (2.45)	rpl18 (0.41)	GAPDH (0.27)	ef1-alpha (0.060)	ef1-alpha (0.34)	GAPDH (2.83)
ef1-beta (0.71)	act (0.89)	GAPDH (0.218)	GAPDH (1.02)	GAPDH (2.83)	ef1-beta (0.44)	act (0.37)	act (0.140)	GAPDH (0.39)	ef1-alpha (3.41)
act (0.74)		act (0.254)	act (1.16)	act (5.00)	ef1-alpha (0.49)			act (0.51)	act (3.98)

The parameter for each software was standard deviation of the  $C_t$  (SD [ $\pm C_t$ ]) for BestKeeper, expression stability value for NormFinder, M value after stepwise exclusion of the highest M value for geNorm, mean of standard deviation of  $\Delta C_t$  for  $\Delta C_t$  method and geomean of ranking values for RefFinder.

The expression level for AChE in developmental stages, when the AChE data were normalized considering the two top ranked gene, enhance from larva to worker (Fig. 5a). AChE from *Anopheles gambiae* also showed a similar expression pattern and some works have shown

that this enzyme also exhibit a noncholinergic functions associated with insect development [43, 44, 45]. In tissues, there is a small variation in the expression of AChE in the head compared to the mesosoma, while in worker without gaster, that is the junction of the other two, the expression is almost 2-fold higher than in mesosoma (Fig. 5b). In *A. gambiae* a higher AChE2 expression was observed in abdomen than in head [45]. Until now, none classification has been done for AChE from *A. sexdens rubropilosa* and we are working in another analysis for an accurate classification.

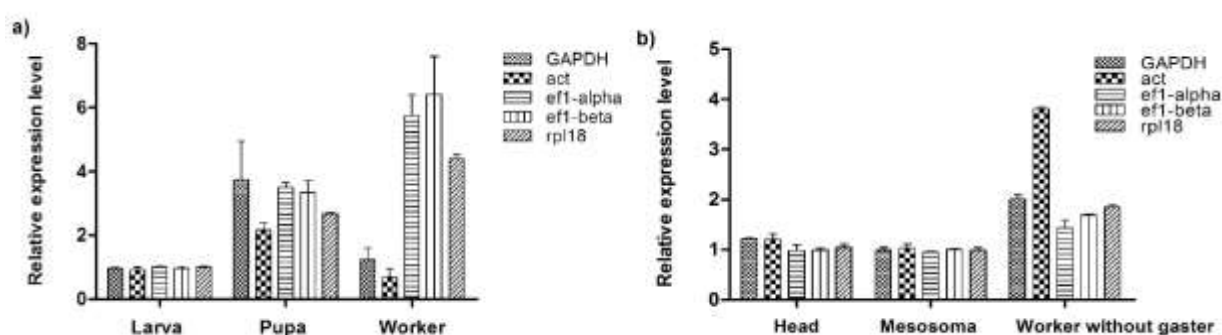


Fig. 5: Relative quantification of AChE in *A. sexdens rubropilosa*. Expression profile of target gene normalized with different candidate reference genes in three A) developmental stages and B) tissues. Data are presented as mean  $\pm$  SEM of biological triplicate.

#### IV. DISCUSSION

RT-qPCR has been widely used for gene expression analysis due to the high accuracy, however, the reliability of the results are strictly correlated with the genes used as reference genes. Due to this, the validation of reference gene is necessary and there are several works suggesting

the importance of the validation for each organism and experimental condition. This is the first work to validate reference gene for *A. sexdens rubropilosa* under biotic condition for different developmental stages and tissues. The results obtained here can support research in this field once leaf-cutting ants is considered as pests in

agricultural crop or reforestation areas mainly in South America.

For the validation for reference gene was done statistical analysis of the data, beside this, different statistical algorithms such as BestKeeper, geNorm, NormFinder, the comparative  $\Delta C_t$  method and RefFinder were used to verify the stability of the genes selected. Once the reference gene can be regulated to some extent, a combination of reference genes should be used, and as indicated by geNorm two reference genes are enough to obtain accurate results. So, we suggest the use of *rpl18* and *ef1-alpha* for developmental stages and *rpl18* and *ef1-beta* for tissues for genomic analysis in *A. sexdens rubropilosa*, based on consensus list provided by RefFinder. For *S. invicta* *rpl18* and elongation factor (*beta*) were the most stable genes for expression in different developmental stages, castes and tissues [12]. The similar results from these two studies was not obvious and experimental results were necessary, once the expression stabilities of HKGs were not conserved among evolutionarily close species [11, 14].

The expression stability values for candidate reference genes are higher for samples from developmental stage than tissue for all algorithm analyzed (Table 3) and this result can be associated with higher complexity of the sample [21]. The transcript profiles from adult stage can change during eclosion process from pupa to adult, as predicted for *S. invicta* resulting in an increase of sample complexity [46]. Moreover, the fact that mature leaf-cutter ant colonies have one of the most complex polymorphic worker caste within ants can contribute to this pattern [24, 47].

Analysis using standard deviation (statistical analysis and BestKeeper algorithm) ranked GAPDH as one of the most stable genes for both development stages and different tissues. All other algorithms listed *rpl18/ef1-alpha* (developmental stages) and *rpl18/ef1-beta* (tissues) as the most stable ones. On the other hand, *act* was classified as the worst for almost all approach. *act* has not been ranked for other Hymenoptera [12, 14], but showed a controversy results for insects from Lepdoptera [11]. This result can be validate by the large number of genes involved in actin cytoskeleton organization identified in *A. cephalotes* compared to other hymenopteran genes that are associated to the extensive cytoskeletal changes that occur during caste differentiation in *Atta* adults [24].

In conclusion, we analyzed five candidate reference genes in two different samples from *A. sexdens rubropilosa* with different statistical approaches, a consensus list from stability of genes was obtained and the two top ranked gene were suggested as reference genes for this insect. The AChE expression pattern normalized with different candidate reference genes emphasize the importance of validation to obtain reliable and accurate results from

gene expression analysis. Beside this, the expression analysis from AChE suggest that this enzyme is important in developmental stage growing from larva to worker and is spread on insect body. The results presented are essential to gene expression analysis in this leaf cutting ant associated with low genome information and the growing interest in pest management control.

#### AUTHORS' CONTRIBUTIONS

DHFS and OCB designed the research and provided guidance; RLC and ACM performed the statistical analysis; ACM and AMS performed the RT-qPCR experiments. DHFS wrote the manuscript. All authors read and approved the final manuscript.

#### ACKNOWLEDGEMENTS

This work was funded by grant 2014/12169-2 from the Sao Paulo Research Foundation (FAPESP).

#### REFERENCES

- [1] H. V. Daly, J. T. Doouen and A. H. Purcell, Introduction to insect biology and diversity, Oxford University Press, New York, 1998.
- [2] AntWeb, Versão 6.3.3., 2016. <http://www.antweb.org> (accessed 16.03.28).
- [3] M. Kaspari and J. D. Majer, Using ants to monitor environmental change, in: Agosti, D., Majer, J.D., Alonso, L.E., Schultz, R., Ants: standard methods for measuring and monitoring biodiversity. Smithsonian Institution Press, Washington, 2000, pp. 89 – 98.
- [4] T. M. C. Della-Lucia, Formigas-cortadeiras: da Biologia ao Manejo, Editora UFV, Viçosa, 2011.
- [5] J. Montoya-Lerma, C. Giraldo-Echeverri, I. Armbrrecht, A. G. Farji-Brener and Z. Calle, "Leaf-cutting ants revisited: towards rational management and control" Int. J. Pest. Manag., vol 53, pp 225–247, 2012.
- [6] F. A. Mariconi, As Saúvas, Agronomica Ceres, Viçosa, 1970.
- [7] E. Hölldobler and E. O. Wilson, The ants, Harvard University Press, Cambridge, 1990.
- [8] J. Huggett, K. Dheda, S. Bustin and A. Zumla, "Real-time RT-PCR normalisation; strategies and considerations" Genes Immun., vol 6, pp 279-284, 2005.
- [9] B. Kozera, and M. Rapacz, "Reference genes in real-time PCR" J. Appl. Genetics, vol 54, pp 391-406, 2013.
- [10] A. Radonić, S. Thulke, I. M. Mackay, O. Landt, W. Siegert and A. Nitsche, "Guideline to reference gene selection for quantitative real-time PCR" Biochem. Biophys. Res. Commun., vol 313, pp 856-862, 2004.

- [11] X. Teng, Z. Zhang, G. He, L. Yang and F. Li, "Validation of reference genes for quantitative expression analysis by realtime RT-PCR in four lepidopteran insects" *Journal of Insect Science*, vol 12, 2012.
- [12] D. Cheng, Z. Zhang, X. He and G. Liang, "Validation of reference genes in *Solenopsis invicta* in different developmental stages, castes and tissues" *Plos one*, vol 8, pp 1-6, 2013.
- [13] W. Fu, W. Xie, Z. Zhang, S. Wang, Q. Wu, Y. Liu, X. Zhou, X. Zhou and Y. Zhang, "Exploring valid reference genes for quantitative real-time PCR analysis in *Plutella xylostella* (Lepidoptera: Plutellidae)" *Int. J. Biol. Sci.*, vol 9, pp 792-802, 2013.
- [14] D. Horňáková, P. Matoušková, J. Kindl, I. Valterová and I. Pichová, "Selection of reference genes for real-time polymerase chain reaction analysis in tissues from *Bombus terrestris* and *Bombus lucorum* of different ages" *Anal. Biochem.*, vol 397, pp 118-120, 2010.
- [15] A. P. Lourenço, A. Mackert, A. S. Cristino and Z. L. P. Simões, "Validation of reference genes for gene expression studies in the honey bee, *Apis mellifera*, by quantitative real-time RT-PCR" *Apidologie*, vol 39, pp 372-385, 2008.
- [16] Y. Lu, M. Yuan, X. Gao, T. Kang, S. Zhan, H. Wan and J. Li, "Identification and validation of reference genes for gene expression analysis using quantitative PCR in *Spodoptera litura* (Lepidoptera: Noctuidae)" *Plos one*, vol 8, 2013.
- [17] F. Ponton, M. P. Chapuis, M. Pernice, G. A. Sword and S. J. Simpson, "Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in *Drosophila melanogaster*" *J. Insect Physiol.*, vol 57, pp 840-850, 2011.
- [18] C. L. Andersen, J. L. Jensen and T. F. Ørntoft, "Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets" *Cancer Research*, vol 64, pp 5245-5250, 2004.
- [19] O. Thellin, W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout and E. Heinen, "Housekeeping genes as internal standards: use and limits" *J. Biotechnol.*, vol 75, pp 291-295, 1999.
- [20] M. Shakeel, X. Zhu, T. Kang, H. Wan and J. Li, "Selection and evaluation of reference genes for quantitative gene expression studies in cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae)" *J. Asia-Pacific Entomol.*, vol 18, pp 123-130, 2015.
- [21] X. Zhu, M. Yuan, M. Shakeel, Y. Zhang, S. Wang, X. Wang, S. Zhan, T. Kang and J. Li, "Selection and evaluation of reference genes for expression analysis using qRT-PCR in the beet armyworm *Spodoptera exigua* (Hübner) (Lepidoptera: Noctuidae)" *Plos one*, vol 9, pp 1-14, 2014.
- [22] S. Guénin, M. Mauriat, J. Pelloux, O. Van Wuytswinkel, C. Bellini and L. Gutierrez, "Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references" *J. Exp. Bot.*, vol 60, pp 487-493, 2009.
- [23] S. Nygaard, G. Zhang, M. Schjøtt, C. Li, Y. Wurm, H. Hu, J. Zhou, L. Ji, F. Qiu, M. Rasmussen, H. Pan, F. Hauser, A. Krogh, C. J. P. Grimmekhuijzen, J. and J. J. Boomsma, "The genome of the leaf-cutting ant *Acromyrmex echinator* suggests key adaptations to advanced social life and fungus farming" *Genome Res.*, vol 21, pp 1339-1348, 2011.
- [24] G. Suen, C. Teiling, L. Li, C. Holt, E. Abouheif, et al., "The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle" *Plos Genet*, vol 7, 2011.
- [25] H. Soreq and S. Seidman, "Acetylcholinesterase - new roles for an old actor" *Nature reviews - Neuroscience*, vol 2, 2001.
- [26] D. C. Harris, *Quantitative Chemical Analysis*, W. H. Freeman and Company, seventh edition, p. 61, 2007.
- [27] R. Bro and A. K. Smilde, "Principal component analysis" *Analytical Methods*, vol 6, pp 2812-2831, 2014.
- [28] K. J. Livak and T. D. Schmittgen, "Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta Ct}$  method". *Methods*, vol 25, pp 402-408, 2001.
- [29] M. W. Pfaffl, A. Tichopad, C. Prgomet and T. P. Neuvians, "Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations" *Biotechnology Letters*, vol 26, pp 509-515, 2004.
- [30] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe and F. Speleman, "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes" *Genome Biology*, vol 3, pp 1-12, 2002.
- [31] N. Silver, S. Best, J. Jiang and S. L. Thein, "Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR" *BMC Molecular Biology*, vol 7, pp 33, 2006.



- [32] F. Xie, P. Xiao, D. Chen, L. Xu and B. Zhang "miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs" *Plant molecular biology*, vol 80, pp 75-84, 2012.
- [33] S. A. Bustin, V. Benes, J. A. Garson, J. Helleman, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele and C. T. Wittwer, "The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments" *Clinical Chemistry*, vol 55, pp 1-12, 2009.
- [34] E. C. Winnebeck, C. D. Millar and G. R. Warman, "Why does insect RNA look degraded?" *Journal of Insect Sci.*, vol 10, pp 159, 2010.
- [35] S. M. Valles, C. A. Strong, E. A. Buss and D. H. Oi, "Non-enzymatic hydrolysis of RNA in workers of the ant *Nylanderia pubens*" *J. Insect Sci.*, vol 12, pp 1-8, 2012.
- [36] S. Fleige and M. W. Pfaffl "RNA integrity and the effect on the real-time qRT-PCR performance" *Mol Aspects Med.*, vol 27, pp 126-139, 2006.
- [37] E. Wang, "RNA amplification for successful gene profiling analysis" *J. Transl. Med.*, vol 3, pp 28, 2005.
- [38] T. Barros Rodrigues, C. Khajuria, H. Wang, N. Matz, D. Cunha Cardoso, F. H. Valicente, X. Zhou and B. Siegfried, "Validation of reference housekeeping genes for gene expression studies in western corn rootworm (*Diabrotica virgifera virgifera*)" *Plos one*, vol 9, 2014.
- [39] R. M. Paim, M. H. Pereira, R. Ponzio, J. O. Rodrigues, A. A. Guarneri, N. F. Gontijo and R. N. Araújo, "Validation of reference genes for expression analysis in the salivary gland and the intestine of *Rhodnius prolixus* (Hemiptera, Reduviidae) under different experimental conditions by quantitative real-time PCR" *BMC Research Notes*, vol 5, pp128, 2012.
- [40] S. Zhang, S. An, Z. Li, F. Wu, Q. Yang, Y. Liu, J. Cao, H. Zhang, Q. Zhang and X. Liu, "Identification and validation of reference genes for normalization of gene expression analysis using qRT-PCR in *Helicoverpa armigera* (Lepidoptera: Noctuidae)" *Gene*, vol 555, pp 393-402, 2015.
- [41] J. Vandesompele, A. De Paepe and F. Speleman, "Elimination of primer-dimer artifacts and genomic coamplification using a two-step SYBR green I real-time RT-PCR" *Anal. Biochem.*, vol 303, pp 95-98, 2002.
- [42] G. W. Takle, I. K. Toth and M. B. Brurberg, "Evaluation of reference genes for real-time RT-PCR expression studies in the plant pathogen *Pectobacterium atrosepticum*" *BMC Plant Biol.*, vol 7, pp 50, 2007.
- [43] G. He, Y. Sun and F. Li, "RNA interference of two acetylcholinesterase genes in *Plutella xylostella* reveals their different functions" *Archives of insect biochemistry and physiology*, vol 79, pp 75-86, 2012.
- [44] M. Kumar, G. P. Gupta and M. V. Rajam, "Silencing of acetylcholinesterase gene of *Helicoverpa armigera* by siRNA affects larval growth and its life cycle" *J. Insect Physiol.*, vol 55, pp 273-278, 2009.
- [45] P. Zhao, Y. Wang and H. Jiang, "Biochemical properties, expression profiles, and tissue localization of orthologous acetylcholinesterase-2 in the mosquito, *Anopheles gambiae*" *Insect Biochem. Mol. Biol.*, vol 43, pp 260-271, 2013.
- [46] L. Ometto, D. Shoemaker, K. G. Ross and L. Keller, "Evolution of Gene Expression in Fire Ants: The Effects of Developmental Stage, Caste, and Species" *Molecular Biology and Evolution*, vol 28, pp 1381-1392, 2011.
- [47] E. O. Wilson, "Caste and Division of Labor in Leaf-Cutter Ants (Hymenoptera: Formicidae: Atta): I. The Overall Pattern in *A. sexdens*". *Behav. Ecol. Sociobiol.*, vol 7, pp 143-156, 1980.